Strategic Management Journal
Strat. Mgmt. J., **20**: 385–395 (1999)

RESEARCH NOTES AND COMMUNICATIONS

ORGANIZATIONAL CONFIGURATIONS AND PERFORMANCE: THE ROLE OF STATISTICAL POWER IN EXTANT RESEARCH

TAMELA D. FERGUSON* AND DAVID J. KETCHEN, JR.
E. J. Ourso College of Business Administration, Louisiana State University, Baton Rouge, Louisiana, U.S.A.

The relationship between organizational configurations and performance has been a frequent albeit controversial subject of research in the field of strategic management. Many studies have failed to find a link between configurations and performance, leading prominent researchers to question the value of the concept of organizational configurations. Before the concept can be discarded, however, other plausible explanations for the lack of findings should be examined. This paper examines the possible role of statistical power. Specifically, it may be that the sample sizes in many studies are too small to detect relationships between configurations and performance when such links are, in fact, present. Analysis of 24 tests of the configurations–performance link revealed that only 8 percent had samples large enough to detect all important relationships. Thus, there is reason to suspect that insufficient statistical power may help to explain extant results. Given these findings, suggestions are presented for improving the ability of configurational studies to detect relationships.

Organizational configurations are sets of organizations that share a common profile along important characteristics such as strategy, structure, and decision processes (Miller and Mintzberg, 1983). For example, Miles and Snow (1978) described firms labeled ‘defenders’ which tend to have narrow market domains, centralized structures, and centralized decision-making. Another set, labeled ‘prospectors,’ operate in broad domains and decentralize their structure and decision-making. Research on configurations encompasses a variety of research streams, including strategic groups (e.g., Hatten and Schendel, 1977), typologies (e.g., Miles and Snow, 1978), taxonomies (e.g., Galbraith and Schendel, 1983), and archetypes (e.g., Miller and Friesen, 1978).

Some configurational research has a singular focus on identifying sets of firms, but studies often also examine if configurations differ along one or more dependent variables. The assumption embedded in the latter approach is that the best way to advance knowledge about organizations is to identify configurations and examine their relations with dependent variables rather than seeking relations that hold across all firms. Organizational performance is often the dependent variable in strategy research (e.g., Rumelt, Schen-

Key words: organizational configurations; statistical power; performance; strategic groups
* Correspondence to: T. D. Ferguson, Ourso College of Business Administration, Rucks Department of Management, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

CCC 0143–2095/99/040385–11 $17.50

Received 18 January 1997
Final revision received 22 June 1998

del, and Teece, 1994), thus it is not surprising that the relationship between configurations and performance has been a frequent research focus.

Whereas the traditional industrial organization economics rationale for expecting that some configurations will outperform others has been criticized (Barney and Hoskisson, 1990), structural contingency theory appears to provide a sound logic (Ketchen, Thomas, and Snow, 1993). Briefly, structural contingency theory asserts that the success of different types of organizations depends on their appropriateness for the environment they face. Configurations that match the requirements of the environment should enjoy more success than configurations that do not. Poor-performing firms would prefer to switch to a configuration that better fits the environment, but such movement is rare because the associated strategic and structural changes are time consuming and expensive (Caves and Porter, 1977). As a result, disparities in fit with the environment should persist and performance differences between configurations can logically be expected (Ketchen *et al.*., 1993).

Despite this logic, empirical results have been mixed. Some researchers have found support for the configurations–performance relationship (e.g., Hawes and Crittenden, 1984; Oster, 1982); others report no connection (e.g., Dowling and Ruefli, 1992; Porter, 1979). This equivocality has created concern about the appropriateness of future inquiry. Indeed, in reference to the most prominent approach to configurations—strategic groups—Barney and Hoskisson (1990: 187) suggest that ‘it may be necessary to abandon this concept’ and redirect attention toward other potential determinants of performance.

Before research on configurations and performance is abandoned, plausible alternative explanations for the lack of findings should be examined. The role of statistical power in extant research provides one such alternative. Statistical power is, in essence, the probability that an empirical test will detect a relationship when a relationship, in fact, exists. Sample size is directly related to power; ceteris paribus, the bigger a sample, the higher the statistical power. Relatively small sample sizes are often found in studies of the configurations–performance link (e.g., 16 in Lewis and Thomas, 1990; 18 in Reger and Huff, 1993). If statistical power is low, this does not necessarily mean that undetected relationships exist, but does indicate that the research is unlikely to find such links if they exist. Given that there is theoretical rationale for expecting performance differences, an assessment of power in extant research is warranted.

Accordingly, the research question addressed by this paper is: Has the statistical power of extant configurations–performance research been strong enough to detect important relationships that may have been present? In order to answer this question, we took the following approach. First, we describe the components of statistical power: effect size, significance level (α), and sample size. The treatment of statistical power in strategy research in general is then briefly reviewed. Next, we assess the statistical power of 24 tests of the configurations–performance relationship published in *Academy of Management Journal* (*AMJ*), *Administrative Science Quarterly* (*ASQ*), *Journal of Management* (*JM*), *Journal of Management Studies* (*JMS*), *Management Science* (*MS*), and *Strategic Management Journal* (*SMJ*). Our main finding is that very few studies have had adequate statistical power. The implications for interpreting the results of past studies are discussed. The paper concludes with suggestions for improving future research.

COMPONENTS OF STATISTICAL POWER

Most empirical research in strategic management is conducted with the hope of rejecting the null hypothesis H_0 (the phenomenon under investigation does not exist, or there is no significant difference between the parameters being tested) in favor of the alternative hypothesis H_a (the phenomenon does exist, or there is a difference in parameters being tested). In this notation, power is the probability that H_0 will be rejected in favor of H_a if H_a is in fact true. More simply, power is the probability of rejecting a false null hypothesis. Power is often represented as $(1-\beta)$, where β is the probability that a type II error (accepting a false null hypothesis) will be made.

Statistical power is determined by three interacting components: effect size, significance level (α), and sample size. Effect size is the proportion of explained variance, i.e., the extent to which variables—such as configuration membership and performance—are related. α is the level of risk of making a type I error (rejecting a true null hypothesis) or 'significance level' that the researcher deems acceptable. The sample size is the number of observations used in a test. If any three of the four elements (power, effect size, significance level, and sample size) are known, the fourth can be calculated, allowing researchers to plan studies based on available information and needs. The following discussion details the characteristics and influence of the statistical power components: effect size, α, and sample size.

Effect size

Effect size describes the magnitude of the relationship between two variables present in a population (Cohen, 1977). When a sample is drawn from a population, a statistical test using the sample provides an estimate of the effect. The effect size for a true null hypothesis is zero; in other words, no relationship exists. Failure to reject the null hypothesis indicates some level of relationship is likely present, with the probability of this relationship emerging by chance expressed as α. Ceteris paribus, the larger the effect size, the more likely it is that a test will detect it. Effect size is frequently expressed in published research as the proportion of explained variance.

The anticipated effect size in a relationship under study should be estimated *a priori* based on meta-analytic calculations, rule-of-thumb estimates, or, preferably, theory (Cohen, 1988). When effects cannot reasonably be estimated, Cohen (1988) suggests using small, medium, and large labels which correspond to 0.20, 0.50, and 0.80 respectively for measures of d, the effect size index for t-tests of means in standard units. The appropriate levels associated with the small, medium, and large labels vary based on the statistical test used. In an F-test, for example, measures of small, medium, and large effect size are 0.10, 0.25, and 0.40 because $d = 2F$. Cohen (1988) provides effect size level definitions for many common statistical tests, including correlation (r), chi-squared tests (w), multiple regression (r^2), and differences between correlations (q). While these levels are simply rules of thumb, Mone, Mueller, and Mauland (1996) found 90 percent of researchers who used power analysis believed 0.80 or greater is correct for t-tests, suggesting widespread acceptance of Cohen's labels.

Alpha (α)

α is the risk of committing a Type I error, i.e., incorrectly rejecting a true null hypothesis. α and statistical power are directly related: the lower the α, the lower the statistical power and the higher the probability of accepting a false null hypothesis. Because an inverse relationship exists between α and β (the probability of a Type II error), one way to increase statistical power is to increase α, thereby reducing β. Researchers must recognize that they determine the level of acceptable risk of both Type I and Type II errors when conducting a test. Specifically, researchers need to be aware of the trade-off involved: as one error risk falls, the other rises.

There are no right or wrong levels of α; thus the choice should be considered as one of several research parameters, with researchers able to choose the level they believe appropriate given other study design elements (Sauley and Bedeian, 1989). The state of theory development is a key contingency. Relatively high significance levels (e.g., $\alpha = 0.10$) may be appropriate when theory about a phenomenon is not developed enough to permit a precise test, whereas lower levels (e.g., 0.01) are desirable when a hypothesis challenges an established body of knowledge.

Although researchers have the choice of significance levels, $\alpha = 0.05$ is by far the most popular. The history of 0.05 can be traced to the work of Fischer in the 1920s. Originally selecting 1 in 20 (i.e., 0.05) as a 'convenient' point to determine significance (Fischer, 1925: 46), he later acknowledged that 0.05 was a personal preference, and that other significance levels were acceptable (Fischer, 1926). By the 1960s, however, the use of 0.05 was embedded in the practice of science to such an extent that authors such as Winer (1962) reminded researchers that this particular level was a matter of convention, not logic. In retrospect, the widespread use of the 0.05 level over the last 50 years appears to be a case of what Popper (1959) labeled conventionalism—the tendency of scientists to agree on issues rather than debate them and then act as if what they have agreed upon is scientific fact. The actual foundation for the use of 0.05 is not logic, but simply an *ad populum* argument: because we all believe, it must be true.

Sample size

In organizational research, sample size may be limited for several reasons, including budget restrictions, the use of convenience samples, interest in small populations, a narrow research focus, and difficulty in obtaining nonpublic or proprietary data (Mone *et al.*, 1996). According to Cohen (1988), even when sample size limitations severely attenuate the power of a test, researchers often prefer to conduct a study, hope for significant results, and then omit reference to the power of the test rather than wait until power can be increased. There are costs to this approach; small samples may foster sample bias and inhibit the generalizability of results. Perhaps most troubling is the fact that null results may be a function of insufficient sample size but might be attributed to theoretical or measurement deficiencies.

The sample sizes needed to have an 80 percent chance of detecting small, medium, and large effects (i.e., 0.80 power) when using an *F*-test are presented in Table 1. If, for example, a researcher expects a small effect (i.e., only a small portion of variance in the dependent variable will be explained by the independent variable), is comparing four groups of observations (e.g., four strategic groups), and adopts $\alpha = 0.05$, then a sample size of 274 is needed. Changes in these parameters dictate change in sample size requirements. If $\alpha = 0.01$ is preferred, for example, the necessary sample size rises to 388.

In sum, statistical power, composed of effect size (the percentage of explained variance), α (the acceptable level of type I error risk), and sample size (the number of observations), is a vital consideration, most appropriately assessed *a priori*. Unfortunately, power is seldom part of research planning in the social sciences (Mone *et al.*, 1996). Many researchers design studies by determining the appropriate statistical test and then simply accepting the common $\alpha = 0.05$, which implies power of 0.80 if the sample size is adequate (see Cohen, 1988: 53–56, for a detailed explanation). Hence, the level of statistical power often is driven by serendipity rather than by planning.

STATISTICAL POWER IN STRATEGIC MANAGEMENT RESEARCH

Most effects in the social sciences are small effects, especially as a field matures and more becomes known about the phenomena under investigation (Cohen, 1977). Debate has raged as to whether strategic management has adopted a paradigm (e.g., Bettis, 1991; Daft and Buenger, 1990), but at a minimum it is clear that theory has advanced significantly and that there are widely shared research questions (Rumelt *et al.*, 1994). Thus, overall, major strides toward maturity have been taken. One implication is that the field of strategic management has developed to the point that the contributions of small effects should be acknowledged.

In the past, the treatment of statistical power in the field of strategic management has evoked criticism. Mazen, Hemmasi, and Lewis (1987) concluded that power was problematic after their analysis of 28 studies published from 1982 to 1984 in *SMJ* and 16 from 1984 in *AMJ* revealed

Table 1. Sample size needed to detect effects at 0.80 power of *F*-test

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
Groups	Sm.	Med.	Lg.	Sm.	Med.	Lg.	Sm.	Med.	Lg.
2	586	95	38	393	64	26	310	50	20
3	464	76	30	322	52	21	258	41	17
4	388	63	25	274	45	18	221	36	15
5	336	55	22	240	39	16	193	32	13
6	299	49	20	215	35	14	174	28	12
7	271	44	18	195	32	13	159	26	11

Adapted from Cohen, J. 1992. 'A power primer', *Psychological Bulletin*, **112** (1), pp. 155–159. Copyright 1992 by the American Psychological Association. Adapted with permission.

that only 23 percent of the studies had adequate power (0.80 or better) to detect small effect sizes, 59 percent could detect medium effects, and 83 percent could detect large effects. Schwenk and Dalton (1991) assessed 77 empirical articles published in 1986 and 1987 in *AMJ*, *ASQ*, *JM*, *MS* and *SMJ*. Only 7 percent had sufficient power to detect small effects, less than one-third could detect medium effects, and two-thirds could detect large effects. Given these results, statistical power was viewed as a major concern that subsequent research should address properly. A recent review by Mone *et al.* (1996) suggests that Schwenk and Dalton's (1991) call has not been heeded. In 30 randomly selected *SMJ* articles published from 1992 to 1994, 18 percent had enough power to detect small effects, 63 percent could detect medium effects, and 87 percent could detect large effects. In both Mazen *et al.* (1987) and Mone *et al.* (1996) nondirectional ('2-tailed') null hypothesis testing at $\alpha = 0.05$ was the basis for power calculations. These authors offered no rationale; perhaps they simply relied, as many have, on 0.05 as a default. The third set of authors, Schwenk and Dalton (1991), was silent on the issue of significance level.

In sum, strategic management studies examined by previous authors were seldom statistically powerful enough to detect all important effects. Given that past strategy research in general has been characterized by low statistical power, assessment of power in organizational configurations–performance studies is logical in order to more precisely understand the state of affairs in the research stream. This assessment is not intended to scrutinize individual researchers or their work, but rather to examine the state of the research stream as a whole.

METHOD

We evaluated the statistical power of all primary studies of the configurations–performance relationship published between 1977, the year the first published strategic groups study appeared (Hatten and Schendel, 1977), and August 1996 in 16 journals comprising the 'forum for strategy research' (MacMillan, 1991). The 11 journals that publish empirical research—*AMJ*, *ASQ*, *Decision Science*, *Journal of General Management*, *Journal of International Business Studies*, *JM*, *JMS*, *MS*, *Omega*, *Rand Journal of Economics*, and *SMJ*—were searched for configurations–performance studies. A total of 24 tests in 18 articles were identified. Eleven of the articles (61%) were published in *SMJ*, three in *AMJ* (17%), and one each in *ASQ*, *JM*, *JMS*, and *MS*.

In past examinations of the power of a body of research (e.g., Mazen *et al.*, 1987), the unit of analysis often has been the article. However, several articles on configurations present multiple, distinct tests. Part of the value of such articles arises from the comparison of the test results, but each test also can independently add to knowledge (e.g., Fiegenbaum and Thomas, 1990). Thus, in this study, the unit of analysis was the statistical test.

For each statistical test, we calculated Eta, the percentage of variance in the dependent variable (performance) associated with membership in a particular group (Cohen, 1977: 282). Eta-squared is literally a generalization of r^2, the most common measurement of effect size, and is appropriate here because more than two configurations are compared in most studies. r^2 is only appropriate when there are two groups. Eta is a nonstandard measure of effect size; therefore it was necessary to transform Eta to f, a standardized measure of effect size associated with the F-test in the analysis (Cohen, 1977). If the configurations examined in a test do not differ in performance, f equals zero (i.e., there is no effect of configuration membership on performance).

Turning to other parameters needed to calculate power, the $\alpha = 0.05$ level was chosen due to its common usage throughout the literature. We also calculated power based on $\alpha = 0.10$ to determine the implications of using a more liberal level. The numerator degrees of freedom, which Cohen (1988) labeled u, was calculated as the number of configurations used in a test minus 1. The size of each sample was identified and labeled n. Using f, n, and u, Cohen's (1988: 311–332) power tables for F-tests were consulted to find each test's ability to detect small (0.10), medium (0.25) and large (0.40) effect sizes. These levels are consistent with standard levels of effect size (0.20, 0.50, and 0.80) associated with d, the effect size index for t-tests, because $d = 2f$.

RESULTS

Table 2 lists each study analyzed, *f*, *n*, *u*, and the probability of detecting small, medium, and large effects at both $\alpha = 0.05$ and $\alpha = 0.10$. The ability of research to detect small effects is the most stringent test of power but, as argued above, detection of small effects is needed in strategic management research. As shown in Table 2, the ability of published tests to detect small effects varied widely. The associated probabilities ranged from 0.06 to 0.99.

Table 3 presents the frequency and ascending cumulative percentage distribution of statistical power sorted by effect size. Using Cohen's 0.80 rule of thumb, only 8 percent of the configurations–performance research could detect a small effect, 46 percent (11 studies) could detect a medium effect, and 75 percent (18 studies) could detect a large effect when using $\alpha = 0.05$. Comparing these results to those obtained in three studies of strategy research in general reveals that the configurations literature has had less power than the studies analyzed by Mazen *et al.* (1987) and Mone *et al.* (1996) and similar power to studies examined by Schwenk and Dalton (1991). Use of the more liberal $\alpha = 0.10$ in our calculations made little difference *vis-à-vis* power. As shown in Table 3, across the literature, one additional study (a total of 12 vs. 11 at 0.05) would have had adequate (i.e., 0.80) power to detect medium effects and two additional studies (20 vs. 18) would have had the power to detect large effects at $\alpha = 0.10$. At both 0.05 and 0.10, only two studies had sufficient power to detect small effects. Given that most effects in the social sciences are small (Cohen, 1977), it appears that adoption of 0.10 would have done little to help past configurations–performance studies find any effects that were present. In summary, as in the field of strategy as a whole, power has been insufficient in configurations research.

DISCUSSION

Our research question asked if levels of statistical power present in extant configurations–performance literature were strong enough to detect important relationships that may have been present. The answer is no. Most effects in the social sciences are modest; thus the ability to detect small effects is not only desirable, but essential. When setting α at 0.05, only 8 percent of extant research was powerful enough to detect small effects. Further, 21 of 24 tests had less than a 50 : 50 chance of detecting a small effect. Simply put, if H_0 is in fact false, the researcher would often be better off assessing H_0 based on a coin flip than on the results of a study.

Implications for interpreting past research

If configurations–performance research was at an exploratory stage, our results might not be as discouraging. Although some observers claim that theory development around the configurations–performance relationship remains weak (Peteraf and Shanley, 1997), the research stream is about 20 years old and has produced numerous empirical studies. In analyzing past research, perhaps the total set of studies should be segmented, with some of the earlier, ground-breaking studies held to less stringent standards regarding power than current work. Unfortunately, Table 2 shows that more recent studies have not been more powerful than their predecessors.

Some researchers have discussed terminating configurations–performance studies based in part on the belief that extant research does not reveal any linkage (Barney and Hoskisson, 1990). Our results do not demonstrate that configurations are a significant predictor of performance. Instead, the results show that research has not had adequate power to detect relationships that may have existed. Many authors suggest that the basic logic for expecting a configurations–performance relationship is robust (Ketchen *et al.*, 1993; Mehra, 1996), even though theory development in individual studies is often lacking. It seems unwise to abandon this line of inquiry until a series of statistically powerful studies have been conducted. With this in mind, we offer some guidelines for conducting future studies.

Suggestions for future research

How can researchers ensure that future configurational studies have adequate statistical power? When designing a study, the probability of detecting a small effect should be determined *a priori*. If a proposed design does not offer adequate power, the researcher has several options: increasing the sample size, increasing

Table 2. Organizational configurations–performance studies and their probability of detecting effect sizes

						Probability of detecting ($\alpha = 0.05$)			Probability of detecting ($\alpha = 0.10$)		
Authors and year	Basis of configuration	No. of configurations	u	n	f	Small effect	Medium effect	Large effect	Small effect	Medium effect	Large effect
Hambrick (1983)	Deductive: Miles and Snow	2	1	850	0.1600	0.99	0.99	0.99	0.99	0.99	0.99
Miller (1988)	Inductive	6	5	792	0.1342	0.99	0.99	0.99	0.99	0.99	0.99
Hawes and Crittenden (1984)	Inductive	3	2	181	0.3296	0.54	0.99	0.99	0.67	0.99	0.99
Lawless and Tegarden (1991)	Inductive	3	2	158	0.1260	0.49	0.99	0.99	0.62	0.99	0.99
Lawless and Finch (1989)[a]	Inductive	5	4	108	0.2939	0.43	0.99	0.99	0.56	0.99	0.99
Robinson and Pearce (1988)	Inductive	5	4	97	0.4325	0.39	0.99	0.99	0.52	0.99	0.99
Ketchen, Thomas, and Snow (1993)[b]	Deductive: Zammuto	4	3	69	0.3627	0.25	0.95	0.99	0.37	0.97	0.99
Ketchen, Thomas, and Snow (1993)[b]	Inductive	4	3	69	0.3378	0.25	0.95	0.99	0.37	0.97	0.99
Snow and Hrebiniak (1980)	Deductive: Miles and Snow	2	1	66	0.2704	0.21	0.82	0.99	0.32	0.89	0.99
Lawless and Tegarden (1991)	Inductive	3	2	57	0.4749	0.19	0.83	0.99	0.31	0.90	0.99
Lawless, Bergh, and Wilsted (1989)	Inductive	2	1	49	0.3896	0.16	0.70	0.97	0.26	0.80	0.99
Mehra (1996)[2]	Inductive	5	4	45	0.6800	0.18	0.86	0.99	0.29	0.92	0.99
Mascarenhas and Aaker (1989)	Deductive: mobility barriers	3	2	33	1.0872	0.13	0.59	0.95	0.22	0.71	0.98
Fiegenbaum and Thomas (1990)	Inductive	3	2	28	1.0256	0.11	0.52	0.91	0.20	0.65	0.95
Cool and Schendel (1987)	Inductive	5	4	20	0.7840	0.10	0.47	0.90	0.18	0.61	0.94
Dess and Davis (1984)	Deductive: Porter	4	3	19	0.5267	0.09	0.41	0.83	0.17	0.54	0.90
Dess and Davis (1984)	Deductive: Porter	3	2	19	0.5152	0.09	0.36	0.76	0.17	0.50	0.85
Tallman (1991)	Inductive	3	2	16	0.3214	0.08	0.31	0.67	0.16	0.44	0.79
Lewis and Thomas (1990)	Inductive	7	6	16	0.8049	0.10	0.45	0.89	0.17	0.58	0.94
Lewis and Thomas (1990)	Inductive	7	6	16	0.7944	0.10	0.45	0.89	0.17	0.58	0.94
Lewis and Thomas (1990)	Inductive	3	2	16	0.5181	0.08	0.31	0.67	0.16	0.44	0.79
Lawless and Finch (1989)	Inductive	3	2	12	0.8091	0.07	0.23	0.53	0.14	0.36	0.67
Lawless and Finch (1989)	Inductive	4	3	7	1.2107	0.06	0.15	0.35	0.12	0.26	0.49

[a]Authors discuss multiple tests of significance, but do not report all results. Only tests with reported results are included.
[b]Number of configurations and (u) represent averages from a series of repeated tests over time.

Table 3. Frequency and ascending cumulative percentage distribution of statistical power

	Small		Medium		Large	
Power	Frequency	Cumulative percent	Frequency	Cumulative percent	Frequency	Cumulative percent
(α = 0.05)						
0.99–1.00	2	8.3%	6	25.0%	11	45.8%
0.95–0.98	0	8.3%	2	33.3%	2	54.2%
0.90–0.94	0	8.3%	0	33.3%	2	62.5%
0.80–0.89	0	8.3%	3	45.8%	3	75.0%
0.70–0.79	0	8.3%	0	45.8%	2	83.3%
0.60–0.69	0	8.3%	1	50.0%	2	91.7%
0.50–0.59	1	12.5%	2	58.3%	1	95.8%
0.40–0.49	2	20.8%	4	75.0%	0	95.8%
0.30–0.39	1	25.0%	4	91.7%	1	100.0%
0.20–0.29	3	37.5%	1	95.8%	0	100.0%
0.10–0.19	8	70.8%	1	100.0%	0	100.0%
0.00–0.09	7	100.0%	0	100.0%	0	100.0%
(α = 0.10)						
0.99–1.00	2	8.3%	6	25.0%	12	50.0%
0.95–0.98	0	8.3%	2	33.3%	2	58.3%
0.90–0.94	0	8.3%	2	41.7%	4	75.0%
0.80–0.89	0	8.3%	2	50.0%	2	83.3%
0.70–0.79	0	8.3%	1	54.2%	2	91.7%
0.60–0.69	2	16.7%	2	62.5%	1	95.8%
0.50–0.59	2	25.0%	4	79.2%	0	95.8%
0.40–0.49	0	25.0%	3	91.7%	1	100.0%
0.30–0.39	4	41.7%	1	95.8%	0	100.0%
0.20–0.29	4	58.3%	1	100.0%	0	100.0%
0.10–0.19	10	100.0%	0	100.0%	0	100.0%
0.00–0.09	0	100.0%	0	100.0%	0	100.0%

the number of configurations, increasing α, or postponing the research until adequate levels of effect size detection can be achieved. Some of these options are methodologically, theoretically, and practically sound whereas others are not.

Perhaps the simplest response to a power shortage is to increase sample size, but many studies have a limited population of interest. For example, Reger and Huff (1993) focused on bank holding companies in one city because their design required interviewees to be intimately familiar with members of the sample. This limited the sample size to 18, but expanding the sample's geographic scope would have precluded the necessary familiarity. Researchers should expand a sample only if expansion does not mortally weaken other aspects of the design.

A second way to increase power is to increase the number of configurations. An overall test of the configurations–performance relationship poses a null hypothesis that states, in essence, 'there are no differences in performance across configurations.' If there are two configurations found in a data set, only one pairwise comparison is made, whereas if four configurations are found each group is compared with three others. With a greater number of comparisons, there is a much better chance of finding substantive differences. For example, as shown in Table 1, if a sample size is 274 and $\alpha = 0.05$ is adopted, a test would not have power of 0.80 if there are two configurations but would if there are four. When identifying the number of groups in a sample, however, researchers attempt to provide an accurate representation of reality. It would be unwise to shift the number of configurations just to obtain additional power.

The relationship between the number of configurations and power has implications for designing research. There is an inverse relationship between the number of configurations in a particular test and sample size requirements; given that power increases as the number of configurations rises, a lower sample size is necessary for the same level of power. Configurational research is unique in that the number of groups of observations that will form the basis for a statistical test is often not known in advance, especially if groups are identified inductively. The researcher must, however, estimate the number of groups that will result in order to use Table 1 to choose an adequate sample size. Extant literature is silent as to how to approach this situation, but a rule of thumb of four groups might be appropriate. This is consistent with two oft-cited approaches to configurations: Miles and Snow's (1978) typology and Porter's (1980) generic strategies. As shown in Table 2, several studies have found four groups, suggesting that four may be a reasonable estimate. Some researchers may wish to be conservative, however, designing a study as if only two groups will be found. Assuming that the other design parameters are addressed correctly, this approach ensures that the sample size provides adequate statistical power, regardless of the number of configurations that emerges.

Increasing the significance level at which the null hypothesis is rejected also enhances power. Researchers should actively contemplate if a level of 0.05 is appropriate for each piece of research. Although the literature as a whole cannot be considered exploratory, studies focusing on unexamined aspects of configurations (e.g., the notion of strategic group identity posited by Peteraf and Shanley, 1997) might be. For an exploratory study, the less conservative level of 0.10 may be appropriate (Sauley and Bedeian, 1989). Increasing α to 0.10 can drastically reduce the sample size needed to have adequate power. As shown in Table 1, obtaining an 80 percent chance of detecting a small effect when there are three configurations requires 322 observations if α = 0.05 but only 258 if α = 0.10: a reduction of 20 percent.

It is important to note that some authors, especially in the field of psychology, are concerned about the value of significance tests regardless of the alpha level selected, arguing that the binary decision-making inherent in such tests (i.e., a null is judged either true or false) is not suitable for many complex behavioral research problems (e.g., Kirk, 1996). To these observers, confidence in one's conclusions lies along a continuum of high to low uncertainty; this variability should be reflected in how studies are reported. To respond to these concerns, configurational researchers might want to report the point estimate of effect size, as well as an appropriate confidence interval, perhaps 95 percent (Hunter, 1997). To date, Lawless, Bergh, and Wilsted (1989) is the only configurational study to report confidence intervals. These authors used the intervals not to examine the configurations–performance link, but rather to determine if firms in the same strategic group had identical capabilities, as determined by those that fell within a 95 percent interval. Future studies could provide more fine-grained information about the configurations–performance link than have past studies by reporting confidence intervals.

If sample size, the number of configurations, and the significance level cannot be changed, a final option is to postpone research until a design with adequate power can be developed. For many researchers, this alternative is unrealistic, given well-known pressures to amass lengthy publication records in order to achieve tenure and/or earn increased compensation (Gomez-Mejia and Balkin, 1992). With this in mind, it may fall to journal reviewers and editors to ensure that power levels are adequate in published research. Indeed, a call for gatekeepers to be sensitive to the importance of power and to require that authors report effect size detection levels seems timely.

Overall, while there are always trade-offs in research design, a more meticulous approach to planning statistical power will be rewarded with increased ability to detect all important relationships that may exist. This guideline applies not only to studies of configurations and performance, but also to strategic management research in general.

ACKNOWLEDGEMENTS

We would like to thank Arthur G. Bedeian, William Ferguson, Anita Heck, Craig Russell, and two anonymous reviewers for their insights on this article.

REFERENCES

Barney, J. B. and R. E. Hoskisson (1990). 'Strategic groups: Untested assertions and research positions', *Managerial and Decision Economics*, **11**, pp. 187–198.

Bettis, R. A. (1991). 'Strategic management and the straight jacket: An editorial essay', *Organizational Science*, **2**, pp. 315–319.

Caves, R. and M. Porter (1977). 'From entry barriers to mobility barriers: Conjectural decisions and deterrence to new competition', *Quarterly Journal of Economics*, **91**, pp. 241–261.

Cohen, J. (1977). *Statistical Power for the Behavioral Sciences* (revised edition). Academic Press, New York.

Cohen, J. (1988). *Statistical Power for the Behavioral Sciences* (2nd ed.). Erlbaum, Hillsdale, NJ.

Cohen, J. (1992). 'A power primer', *Psychological Bulletin*, **112**(1), pp. 155–159.

Cool, K. O. and D. Schendel (1987). 'Strategic group formation and performance: The case of the U.S. pharmaceutical industry, 1963–1982', *Management Science*, **33**, pp. 1102–1124.

Daft, R. L. and V. Buenger (1990). 'Hitching a ride on a fast train to nowhere: The past and future of strategic management research', *Perspectives in Strategic Management*, pp. 81–103.

Dess, G. G. and P. S. Davis (1984). 'Porter's (1980) generic strategies as determinants of strategic group membership and organizational performance', *Academy of Management Journal*, **27**, pp. 467–488.

Dowling, M. J. and T. W. Ruefli (1992). 'Technological innovations as a gateway to entry: The case of the telecommunications equipment industry', *Research Policy*, **21**, pp. 63–77.

Fiegenbaum, A. and H. Thomas (1990). 'Strategic groups and performance: The U.S. insurance industry, 1970–1984', *Strategic Management Journal*, **11**(3), pp. 197–215.

Fischer, R. A. (1925). *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.

Fischer, R. A. (1926). 'The arrangement of field experiments', *Journal of the Ministry of Agriculture*, **33**, pp. 503–513.

Galbraith, C. and D. Schendel (1983). 'An empirical analysis of strategic types', *Strategic Management Journal*, **4**(2), pp. 153–173.

Gomez-Mejia, L. R. and D. B. Balkin (1992). 'Determinants of faculty pay: An agency theory perspective', *Academy of Management Journal*, **35**(5), pp. 921–955.

Hambrick, D. C. (1983). 'Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types', *Academy of Management Journal*, **26**, pp. 5–26.

Hatten, K. J. and D. E. Schendel (1977). 'Heterogeneity within an industry', *Journal of Industrial Economics*, **26**, pp. 97–113.

Hawes, J. M. and W. F. Crittenden (1984). 'A taxonomy of competitive retail strategies', *Strategic Management Journal*, **5**(3), pp. 275–287.

Hunter, J. E. (1997). 'Needed: A ban on the significance test', *Psychological Science*, **8**(1), pp. 3–7.

Ketchen, D. J., Jr., J. B. Thomas and C. C. Snow (1993). 'Organizational configurations and performance: A comparison of theoretical approaches', *Academy of Management Journal*, **36**, pp. 1278–1313.

Kirk, R. E. (1996). 'Practical significance: A concept whose time has come', *Educational and Psychological Measurement*, **56**(5), pp. 746–759.

Lawless, M. W., D. D. Bergh and W. D. Wilsted (1989). 'Performance variations among strategic group members: An examination of individual firm capability', *Journal of Management*, **15**, pp. 619–661.

Lawless, M. W. and L. F. Tegarden (1991). 'A test of performance similarity among strategic group members in conforming and non-conforming industry structures', *Journal of Management Studies*, **28**, pp. 645–664.

Lawless, M. W. and L. K. Finch (1989). 'Choice and determinism: A test of Hrebiniak and Joyce's framework on strategy–environment fit', *Strategic Management Journal*, **10**(4), pp. 351–365.

Lewis, P. and H. Thomas (1990). 'The linkage between strategy, strategic groups, and performance in the U.K. retail grocery industry', *Strategic Management Journal*, **11**(5), pp. 385–397.

MacMillan, I. C. (1991). 'The emerging forum for business policy scholars', *Strategic Management Journal*, **12**(2), pp. 161–165.

Mascarenhas, B. and D. A. Aaker (1989). 'Mobility barriers and strategic groups', *Strategic Management Journal*, **10**(5), pp. 475–485.

Mazen, A. M., M. Hemmasi and M. F. Lewis (1987). 'Assessment of statistical power in contemporary strategy research', *Strategic Management Journal*, **8**(4), pp. 403–410.

Mehra, A. (1996). 'Resource and market based determinants of performance in the U.S. banking industry', *Strategic Management Journal*, **17**(4), pp. 307–322.

Miles, R. E. and C. C. Snow (1978). *Organizational Strategy, Structure, and Process*. McGraw-Hill, New York.

Miller, A. (1988). 'A taxonomy of technical settings with related strategies and performance levels', *Strategic Management Journal*, **9**(3), pp. 239–254.

Miller, D. and P. H. Friesen (1978). 'Archetypes of strategy formulation', *Management Science*, **24**(9), pp. 921–933.

Miller, D. and H. Mintzberg (1983). 'The case for configuration'. In G. Morgan (ed.), *Beyond Method: Strategies for Social Research*. Sage, Newbury Park, CA, pp. 57–73.

Mone, M. A., G. C. Mueller and W. Mauland (1996). 'The perceptions and usage of statistical power in applied psychology and management research', *Personnel Psychology*, **49**, pp. 103–120.

Oster, S. (1982). 'Intraindustry structure and the ease of strategic change', *Review of Economics and Statistics*, **64**, pp. 376–384.

Peteraf, M. and M. Shanley (1997). 'Getting to know

you: A theory of strategic group identity', *Strategic Management Journal*, Summer Special Issue, **8**, pp. 165–186.
Popper, K. R. (1959). *The Logic of Scientific Discovery*. Harper & Row, New York.
Porter, M. E. (1979). 'The structure within industries and companies' performance', *Review of Economics and Statistics*, **61**, pp. 214–227.
Porter, M. E. (1980). *Competitive Strategy*. Free Press, New York.
Reger, R. K. and A. S. Huff (1993). 'Strategic groups: A cognitive perspective', *Strategic Management Journal*, **14**(2), pp. 103–123.
Robinson, R. B., Jr. and J. A. Pearce II (1988). 'Planned patterns of strategic behavior and their relationship to business-unit performance', *Strategic Management Journal*, **9**(1), pp. 43–60.
Rumelt, R. P., D. E. Schendel and D. Teece (1994). *Fundamental Issues in Strategy: A Research Agenda*. Harvard Business School Press, Boston, MA.
Sauley, K. S. and A. G. Bedeian (1989). '.05: A case of the tail wagging the distribution', *Journal of Management*, **15**(2), pp. 335–344.
Schwenk, C. R. and D. R. Dalton (1991). 'The changing shape of strategic management research', In P. Shrivastava, A. Huff and J. Dutton (eds.), *Advances in Strategic Management*, Vol. 7. JAI Press, Greenwich, CT, pp. 277–300.
Snow, C. C. and L. G. Hrebiniak (1980). 'Strategy, distinctive competence, and organizational performance', *Administrative Science Quarterly*, **25**, pp. 317–336.
Tallman, S. B. (1991). 'Strategic management models and resource based strategies among MNEs in a host market', *Strategic Management Journal*, Summer Special Issue, **12**, pp. 69–82.
Winer, B. J. (1962). *Statistical Principles in Experimental Design*. McGraw-Hill, New York.